

REMARKS

The present application was filed on June 23, 2003 with claims 1-22. Claims 1, 10, 19, 21 and 22 are the independent claims.

In the outstanding Office Action, the Examiner rejected claims 1-22 under 35 U.S.C. §102(a) as being anticipated by a Garg et al. article entitled "Frame-Dependent Multi-Stream Reliability Indicators for Audio-Visual Speech Recognition," ICASSP, vol. 1, pp. 24-27, April 2003 (hereinafter "Garg").

In this response, Applicants amend the present specification to correct minor typographical errors, and traverse the §102(a) rejection of claims 1-22, for at least the following reasons.

Regarding the issue of whether claims 1-22 are anticipated under 35 U.S.C. §102(a) by Garg, the Office Action contends that Garg discloses all of the claim limitations recited in the subject claims. Applicants respectfully assert that Garg fails to teach or suggest all of the limitations in claims 1-22, for at least the reasons presented below.

It is well-established law that a claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631, 2 U.S.P.Q.2d 1051, 1053 (Fed. Cir. 1987). Applicants assert that the rejection based on Garg does not meet this basic legal requirement, as will be explained below.

The present invention, for example, as recited in independent claim 1, recites a method for use in accordance with an audio-visual speech recognition system for improving a recognition performance thereof, comprising the steps of selecting between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment, and decoding at least a portion of an input spoken utterance using the selected data model. Independent claims 10, 19 and 21 recite similar limitations.

Advantageously, as illustratively explained in the present specification at page 2, during periods of degraded visual conditions, the audio-visual speech recognition system is able to decode (recognize) input speech data using audio-only data, thus avoiding recognition inaccuracies that may result from performing speech recognition based on acoustic-visual data models and degraded visual data.

Furthermore, as illustratively explained in the present specification at page 2, principles of the invention may be extended to speech recognition systems in general such that model selection (switching) may take place at the frame level. Switching may occur between two or more models. By way of example, independent claim 22 recites a method for use in accordance with a speech recognition system for improving a recognition performance thereof, comprising the steps of selecting for a given frame between a first data model and at least a second data model based on a given condition, and decoding at least a portion of an input spoken utterance for the given frame using the selected data model.

Garg, as explained in its Abstract on page 24, investigates the use of local, frame-dependent reliability indicators of the audio and visual modalities, as a means of estimating stream components of multi-stream hidden Markov models (HMM) for audio-visual speech recognition system.

That is, Garg proposes using soft weights on each of the audio and visual HMM modalities. The value of this weight is determined through a likelihood ratio test based on observations in the acoustic space only. The dispersion metric is based on speech class conditional likelihoods, in this case, speech context dependent or independent phonemes.

However, nowhere in Garg is there any teaching or suggestion of actually selecting between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment, and decoding at least a portion of an input spoken utterance using the selected data model, as recited in independent claims 1, 10, 19 and 21. Nor does Garg teach or suggest selecting for a given frame between a first data model and at least a second data model based on a given condition, and decoding at least a portion of an input spoken utterance for the given frame using the selected data model, as recited in independent claim 22.

That is, the invention provides a selection process whereby the recognition system decides whether or not to incorporate visual information in the decoding process. The decision is based on “a condition associated with the visual environment,” i.e., the decision is derived from observations outside the acoustic space.

In contrast, as pointed out above, Garg determines weights on each of the audio and visual HMM modalities through a likelihood ratio test based on observations in the acoustic space only.

In fact, to further point out the significant differences between in the claimed invention and Garg, it is to be appreciated that after the selection process defined by the claimed invention, the Garg weighting process could be utilized. That is, once a decision was made to utilize visual data in the recognition process, in accordance with selection of the audio-visual model (rather than selection of the audio-only model), the weighting described in Garg could be used.

For at least the above reasons, Applicants assert that independent claims 1, 10, 19, 21 and 22 are patentable over Garg.

Regarding the §102(a) rejection of claims 2-9, 11-18 and 20, it is respectfully submitted that such claims directly or indirectly depend from independent claims 1, 10 and 19 and are therefore patentable for the reasons that claims 1, 10 and 19 are patentable. However, it is also respectfully submitted that said dependent claims are patentable over Garg because they recite patentable subject matter in their own right.

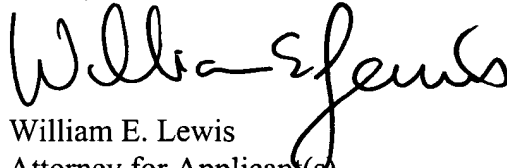
By way of example, dependent claims 2-9, 11-18 and 20 recite limitations pertaining to the model selection step/operation. However, since Garg discloses nothing whatsoever about a model selection step/operation, Garg is also silent regarding the details of a model selection step/operation.

Further, claims 2, 11 and 20 recites storing the acoustic-only data model and the acoustic-visual data model in memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located. Despite the assertion to the contrary in the Office Action, Garg is completely silent to any pointer shifting operation.

The Office Action generally cites the various sections of the Garg paper against the various claimed details, but a review of these sections reveals that they clearly do not teach or suggest what the Examiner suggests they do.

In view of the above, Applicants believe that claims 1-22 are in condition for allowance, and respectfully request withdrawal of the §102(a) rejection.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "William E. Lewis". The signature is fluid and cursive, with the first name "William" being the most prominent part.

Date: September 8, 2005

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946